JPRS: 3629

10 August 1960

SOVIET DEVELOPMENTS IN INFORMATION PROCESSING AND MACHINE TRANSLATION

U. S. JOINT PUBLICATIONS RESEARCH SERVICE
205 EAST 42ND STREET, SUITE 300
NEW YORK 17, N. Y.

F O R E W A R D

This publication was prepared under contract by the UNITED STATES JOINT PUBLICATIONS RESEARCH SERVICE, a federal government organization established to service the translation and research needs of the various government departments.

JPRS: 3629

CSO: 3901-D/25

SOVIET DEVELOPMENTS IN INFORMATION PROCESSING AND MACHINE TRANSLATION

FOREWORD

This translation series presents information from Soviet literature on developments in the following fields in information processing and machine translation: organization, storage and retrieval of information; coding; programming; character and pattern recognition; logical design of information and translation machines; linguistic analysis with machine translation application; mathematical and applied linguistics; machine translation studies. The series is published as an aid to U. S. Government research.

Previously issued JPRS reports on this subject include:

JPRS: 68, 241, 319, 355, 379, 387, 487, 621, 646, 662, 705, 729, 863, 893, 925, 991, 992, 1006, 1029, 1130, 1131, 1132, 1133, 3225, 3356, 3433, 3502, 3532, 3590, 3597, 3598, 3599 and 3613.

SOVIET DEVELOPMENTS IN INFORMATION PROCESSING AND MACHINE TRANSLATION

[Following is a translation of selected articles from the Russian-language periodical Voprosy Yazykoznaniya (Problems in Linguistics), Moscow, No. 2, March/April 1960. Page and name of author, if available, are given under individual article headings.]

Table of Contents

I. THE LONG-TERM PLAN OF SOVIET LINGUISTIC RESEARCH DURING THE NEXT FEW YEARS

Pages 3-10 Unsigned article

The science of language in the Soviet Union is now on the threshold of a new stage in its history. During the past forty years it has trodden a difficult and rough road. On the way it gained a tremendous amount of rich and diversified experience. The articles that appeared in our journal from time to time recorded the state of our linguistics at the various phases of its development and served as signposts of this development. There is no particular point in focussing on the past. Rather, let us look toward the future, but with due regard for the lessons of the past and what we see at present.

1

The theoretical foundation of our linguistics has naturally remained unchanged: we consider language a social phenomenon sui generis. Our own experience as well as observation of the state of linguistic thought in other countries have only confirmed the validity and cogency of our view. This, in turn, conditions everything else: the direction of our work on the general problems of linguistics and our attitude toward the world's current thinking on the subject.

Our view of language as a social phenomenon requires, above all, strict historicity. Linguistic theory must be based on data derived from the history of specific languages differing both in type of structure and in characteristic method of development. Linguistic activity in our country with its numerous languages varying in structure, level of development, and history provides us with the richest of material. We already have an extensive and valuable literature describing the individual languages and groups of languages (e.g., Caucasian, Turkish, Mongolian, Manchuro-Tungusic, Finno-Ugric), including those described for the first time with great care and in abundant detail (e.g., paleoasiatic). The descriptive work must be intensively continued, the research material steadily enlarged. This work is of primary significance for further progress in educating the peoples of the Soviet Union and for cultural construction in general. At the same time, however, we must substantially broaden our work on the history of all these languages, at any rate the history of those languages which are accessible to investigation. It is obvious that such historical investigation must also be extended to the languages of other peoples, especially those which provide new and unusual material for the study of linguistic structure - development, movement, internal transformations.

To the study of the history of languages we must add the study of linguistic thought on the widest scale possible. Every major culture whose language has had a long history reflected in written monuments has its own ideas about the nature, structure, and functions of its language. Some peoples developed their own linguistics long ago, sometimes highly advanced in its practical as well as theoretical aspects. This "native" linguistics arose among the peoples of India, Europe, Arabia, China, and Japan. The separate streams of language science may well have been genetically related, i.e., they may have originated from the same source or influenced one another. This is a fact of the greatest importance and creates the soil for the appearance of some common features. That is why the history of individual languages and groups of languages combined with a study of the history of native linguistic thought in all its manifold relations among the various peoples and a study of the general movement of languages in the history of mankind may serve as a reliable criterion for evaluating modern general linguistic theories. This will be possible, of course, only if such a comprehensive study takes place within the framework of the social and cultural development of each people and of mankind as a whole.

The comprehensive analysis and description of individual languages and groups of languages, history of individual languages and groups of languages, relations between languages, and the linguistic thought of various peoples at the different stages of their history has for its objective the construction of a Marxist theory of linguistics. This also constitutes the major task of our science. This task, of course, cannot be completed within an arbitrary period of time. It is important, however, that all the investigations, even of the apparently most specialized kind, have the same goal, thereby not only guaranteeing its eventual attainment, but also enhancing the scientific value of the individual efforts.

2

We must also study the linguistic activity of our times. It is essential that all linguists, even those specializing in the languages of antiquity, pay scientific attention to actual language usage, which imparts a necessary feeling of reality without which research risks being abstract and scholastic. What is happening now retrospectively illuminates much of what happened in the past and, to some degree, gives an indication of what may happen in the future. Moreover, the language processes now taking place are helpful in enabling us to gain a better understanding of the origin and essence of the different theories that arose in world linguistic science after the period of the

young grammarians, especially during the last twenty or thirty years, and to see therein that which actually reflects language reality and that which is merely ascribed to it or is generally fashioned with no relation whatever to reality.

If we examine modern languages closely, we find them increasingly dividing into spoken and written speech, resulting in the considerable isolation of each as a distinctive form. This division has nothing in common with the earlier division into "colloquial" and "literary" (book) found in the history of many sophisticated languages. At one time this division was largely caused by the inadequate level of development of the standard language, the difference in degree and limits of enlightenment and education among the various layers of the pupulation, and the language policy of the ruling classes. The present division has been caused by the vast development of mass communication and the intense growth of the exact sciences.

Mass communication is a consequence of the extraordinary broadening of social life, of the mounting participation of the masses in the political and cultural life of the country. It is sustained and expanded by the heightened struggle of the people for the right to such participation and by the process of democritizing culture and enlightenment. The growth of the exact sciences is a result of the continuous and steadily increasing demands made by society with its new needs and attraction to the exact sciences.

Since a major feature of mass communication is explanation and persuasion, the main instrument in colloquial language is the energetic mobilization of the expressive resources of language drawn from the sphere of meaning, from the sphere of intonation, and from the sphere of methods of uniting and disuniting the various elements of speech. As colloquial speech is recorded in writing, methods of adequately compensating for the absence of the elements of sound become available. The result has been the recent prominence of those branches of linguistic science whose object is the phenomena of meaning, intonation, and, in general, the sound aspect of speech and syntax. Thus, many new theoretical concepts have been introduced into science that genuinely reflect some part of modern linguistic activity, specifically, the so-called "grammar of expression" (grammatika vyrazheniya).

Information is primarily precise knowledge. Language as a means of transmitting information does not need the expressive resources mentioned above. Instead, it can mobilize with all its power the sound aspect of language, converting thereby a word so to speak into a hieroglyphic symbol and a syntactic construction into a hieroglyphic combination. This has led to the formation of several theoretical concepts which reflect in large

measure another aspect of modern linguistic activity, specifically, the so-called "sign theory" (znakovaya teoriya) of language.

The development of our linguistics requires equally a study of these phenomena of modern linguistic activity and a study of the linguistic theories arising from them. However, this research can be truly fruitful only if there is a clear-cut understanding of social conditionality both of the actual phenomena of the linguistic activity and of all the concepts derived therefrom.

The task of modern linguistics is to study interlingual relations. In this day and age of wide dissemination of the same knowledge, education, enlightenment, and extraordinary development of international life with the criss-crossing, drawing together, and interweaving of people as well as the clash of different tendencies and interests, the people who represent contemporary civilization use more or less the same ideas and concepts in their languages. While each nation naturally renders them in its own words, the content and usage as a whole are generally identical. The concept expressed by the Russian phrase soveshchaniye v verkhakh ("summit meeting") is translated in German by the compound Gipfelkonferenz, somewhat differently in other languages, but in all cases the idea is the same. This phenomenon, of course, is far from being new, it is just that the range is unusually wide. Since every linguistic phenomenon that receives particularly extensive circulation has consequences of significance for language as a whole, the development of an "international lexicon" in a "national form" must be carefully watched. This is necessary because the development of such a lexicon foreshadows the prospect of languages coming together with each preserving its own national material base. The widespread practice of synchronous translation is a laboratory for creating and checking the elements of an international language in different national forms. It is achieving something absolutely new in the sphere of language as a social phenomenon: the fact of simultaneous multilingual communication.

3

The varied problems confronting Soviet linguists during the next few years can, with some over simplification, be reduced to the following four main categories: (1) theory of Soviet linguistics, (2) historical-comparative study of families and groups of languages as well as individual languages and dialects, (3) study of the structure of modern languages and dialects,

(4) fundamental questions in semeiology, information theory, and applied linguistics. We must discuss each briefly in the light of the most general evaluation of the urgent requirements of our times.

The main research task in the field of Soviet linguistic theory is to create a comprehensive Marxist doctrine on language as a social phenomenon and the laws of its historical development. Elaboration of the entire set of theoretical problems of Soviet linguistics, as mentioned above, must be based on concrete historical research on languages differing both in structure and sociohistoric conditions. Plans for the coming decade call for a study of the development and use of languages at various eras from earliest preclassical to modern times. Emphasis will be placed on specific changes in languages and dialects at various stages in the building of socialism and in the period of transition from socialism to communism. A narrow aspect of the problem is to be dealt with in the next few years in working out the topic "Russian language and Soviet society" with subsequent expansion of the material by embracing other languages of the USSR (more than 120, including some 70 literary languages) and then the languages of all the countries in the socialist camp. From historical studies of individual languages and language groups will be derived generalizations on the correlation between the laws of internal development (self-development) of linguistic systems and those changes by which a language directly responds to the new requirements of consciousness and communication.

This research area will include in part communication and the correlation of language and thought, a problem scarcely considered by philosophers, psychologists, or linguists during the past ten years. Only the preparatory work can be done in the near future on the psychology of thought (connection between the processes of thought and speech, "internal speech," two-way nature of the process of oral communication), typology of languages, in particular, translatability of languages, and, finally, experimental language "modeling" (based on the use of mathematical methods) which should provide a more precise idea of the "mechanisms of speech." The problem of "language and thought" is to be worked out along with cybernetic problems, making use of "information theory" in order to study the correlation of structural elements in a "sign system" in the broadest sense of the term whereby the same algorithms will be applied to the most varied systems. Thus, the problem is closely connected with the problem of language as a system and amenable to the methods of structural analysis, i.e., the theoretical problems of structural linguistics, the study of language on the synchronous plane. A variety of research activities are proposed that should result in the creation of a much more precise method

of language description and analysis on this plane. These projects are to be accompanied by a critical study of the research techniques of foreign linguists with a view to determining the possibility of using the valuable features of these techniques. After conclusion of the initial work on the individual "levels" of language as a system (phonological, phonomorphological, morphematic, syntagmatic, etc.), it will be necessary to generalize the structural principles of the system as a whole and its relation to thought.

The literary language, a problem that has been extensively studied by Soviet scholars for the past twenty-five years, occupies a prominent place in the theory of Soviet linguistics. Fiction is now seen in a new light as a problem of language - a means of descriptive expression of reality, i.e., something very specific. The language of business prose, journalism, and science is viewed in another way. The first task is to work out more "objective" methods, precise terminology, and an orderly system of concepts. Individual problems requiring investigation here are: the specific features of poetry, prose, and dramaturgy, language style and speech style, stylistic differentiation of vocabulary, value of realistic fiction for the development of a modern literary language, etc.

A theoretical problem, the solution of which is a prequisite to the solution of any of the problems mentioned above, is that of the aforementioned laws of interaction of languages. The problem, which has scarcely been studied in detail, includes the formation of common structural and conceptual elements in the major languages of the modern world resulting from increasing communication between peoples.

A solution of these problems is inseparable from an analysis of the full set of modern linguistic theories and the scientific legacy of the past, from a logical criticism of non-Marxist theories and identification in the development of foreign linguistic thought of the progressive elements that might be utilized in the forward march of Soviet science.

4

We have already pointed out that the main characteristic of Soviet linguistics is its historism, its striving to discover the laws governing the historical development of languages and dialects. An important role in this effort is played by the comparative-historical study of related languages, whose laws are partly determined by tendencies rooted in deep antiquity when these languages were all alive. The comparative-historical method elaborated during the 19th century and substantially refined in recent decades still occupies a prominent place in Soviet

historical linguistics. This method, however, must be based on a detailed investigation of the history of the peoples and combined with a study of the interaction of languages and the effect of different "substrata" as phenomena arising from the specific history of the peoples; it must also be combined with comparative-typological investigations. The recent introduction into comparative-historical linguistics of the methods of linguistic geography crystallized in the study of living dialects has led to the creation of a special discipline - "areal linguistics," which gives promise of further expanding the field of research. Comparative linguistics is being strengthened by new material from toponymic and onomastic studies.

Comparative linguistics in the Soviet Union continues to lag far behind other fields mainly because insufficient attention is paid to the languages and dialects of many families and groups of languages in the U.S.S.R. A study of the structure of living languages should in the course of time also yield new material for comparative linguistics.

The tasks of comparative-historical research on the languages of such families as the Finno-Ugric, Turkic, Tungusic-Manchu, Caucasian, etc are quite specific, being determined by the absence or scantiness of written monuments of great antiquity and by the existence of numerous dialects, which calls for highly specialized methods of research. There are, moreover, urgent problems in the vocabulary of related languages, i.e., in the field of etymological studies, which likewise requires great precision in technique and the creation of reliable etymological dictionaries.

Another aspect of historical linguistics is the history of literary languages and common national colloquial languages in relation to the history of dialects. This also includes the history of social dialects, argots, and various branches and systems of technical terminology - an area of research that we have almost forgotten in recent times. In the investigation of dialects we must shift from the descriptive monographs prevailing hitherto to the creation of a historical dialectology of individual languages, to a study of the genesis of individual dialectic groups in relation to the historical fate of the people speaking them. Historical dialectology crosses with the investigation of literary languages in the field of research on dialectic varieties of the common national colloquial language and their impact on the literary language. Historical and comparative-historical studies of languages and groups of languages cannot be divorced from, indeed it relies on, research on the structure of modern languages and dialects.

The objective of research here is to apply the principles and methods worked out by the theory of Soviet linguistics to modern languages and dialects (primarily those of the U.S.S.R.) and at the same time to perfect the techniques of describing languages on the synchronous plane, which, in turn, must be utilized for further development of the theory.

Besides describing unstudied or little studied languages of the peoples of the U.S.S.R. and describing in greater detail the major languages (Russian, Ukrainian, Georgian, Armenian, Kazakh, etc), we must investigate the typological characteristics of modern languages, which are of practical importance in machine translation and in elaborating correct principles of language pedagogy. The interrelation between the general and specific features of structural models of different groups of related and unrelated languages must be clarified. Research in this still almost virgin field will have to define more exactly the very concept of "language type." The problem of language typology is closely connected in part with cybernetics and in part with the problem of "language and thought." The translatability of languages has to be dealt with in accordance with the Marxist view concerning the unity of the categories of human thought despite the limitless variety of linguistic expression. We must also stress the importance of comparative-typological investigations dealing with the structural-semantic analysis of individual linguistic units ("word," "word combination," "clause") in different types of languages.

Both series of investigations will have to culminate in a general typology of the languages of the world, although some preliminary generalizations on the languages of the U.S.S.R. may be given (along with the results of the descriptive studies of groups of genealogical classification) in the encyclopedic work Yazyki Narodov SSSR (Languages of the Peoples of the U.S.S.R.) to be issued in connection with the fiftieth anniversary of the Great October Socialist Revolution.

This field of research will also include questions bearing on standardization of the literary languages of the U.S.S.R. Significant progress in this branch of practical linguistics will have to await conclusion of theoretical research seeking to define the concepts "modern language" and "language norm" (in connection with the difference in the principles of standardizing old and young written languages). It is also necessary to investigate the laws and tendencies affecting the appearance of new phenomena in modern languages taking into account their historical traditions and interdialect and interlanguage relations and to study the development of the "linguistic tastes"

of society and the correlation between the conscious influence of society on language and the inherent development of language itself.

5

Progress in modern science is leading, on one hand, to increasing differentiation and systematization of the various disciplines and, on the other, to the rise of new disciplines in allied fields on the borderline between different sciences and to the appearance of hitherto non-existent synthesizing branches of science (e.g., cybernetics). The two-sided approach to language - functional, as a means of communication between peoples, and immanent, as a mechanism serving to effectuate this communication - explains and, to a certain degree, justifies the somewhat tentative division of linguistics into "external" and "internal" linguistics (or, better, into "social" and inherent"). The former is connected with history, archeology, ethnography, and other social sciences, the latter with cybernetics, mathematics, and other exact sciences. There is, however, no sharp dividing line between the two and the question of excluding linguistics from the humanities cannot be seriously entertained. The period when the "new doctrine of language" prevailed was marked by extreme neglect of problems in internal linguistics and the "speech mechanism." That is why Soviet linguistics, which is striving to extract everything that is useful and fruitful from the world arsenal of linguistic science, must now stress semeiology, information theory, and applied linguistics. This trend embraces a number of independent problems that can be solved only if there is close cooperation among mathematicians, physicists, physiologists, and specialists in cybernetics, information theory, electronics, acoustics, etc. The effort must be interdepartmental, and the linguistic institutes of the Academy of Sciences of the U.S.S.R. and republic academies are to work out just a part of the problems and only those dealing with the linguistic aspects. However, if the designated goals are to be reached, personnel who are competent in some branches of the allied sciences will have to be trained in applied linguistics. Therefore, the scale of research, necessarily modest during the next few years, will eventually have to be steadily expanded.

The most important thing here is the use of mathematical methods, which may raise to new heights the accuracy of linguistic analysis and conclusions drawn therefrom. Of importance too are the application of probability analysis, set theory, and mathematical statistics to language phenomena and the investigation of all "levels" of language structure in connection with the general theory of semeiology and "information theory." The work

will have to be performed on the practical as well as theoretical planes - machine translation, determination of phonological variation and phonetic combinations, systematization of grammatical rules for combining morphemes and words into possible units and identification of the classes of these units, and rational rules for planning and compiling different types of dictionaries.

Some of the experimental phonetic research will also be conducted in the light of cybernetics problems: work on the sound elements of language, on the syllable, time, sentence, etc. For practical purposes it will be necessary to study the differential features of phonemes in their combining characteristics, syllable formation, boundary signals, rhythm of different speech segments, intonational structure of the sentence, etc. All this is essential to establish the correlation between different sets of language data and to determine the possibilities of oral input in translation machines, automatic control of mechanisms and processes by means of speech in connection with the possible compression of the physical characteristics of speech (taking into account redundant and adequate information on all the structural levels of language).

The next problem is that of machine translation in its linguistic aspects, which is related both to the general theory of translation and to special "applied" subjects, specifically to research in the field of analysis and synthesis, programming of translation algorithms for different languages with due regard for their structural typology. It is self-evident that intensified efforts will be required to work out interlanguage correspondences, determine the structure of the intermediary language in its various aspects, and create a special language for recording machine translation algorithms.

The field of applied linguistics also includes extension of the principles underlying the creation or regulation of specialized technical terminology, the principles of orthography, transcription and transliteration, and the problem of international auxiliary languages requiring an investigation of the principles governing the creation of artificial languages, including a symbolic language for science as a whole and for the different sciences. This research will be conducted in the light of the general theory of semeiology and codes, taking into account data of the theory of linguistic transformations and conversions in coding and recoding and typology of languages.

Neither the subject matter of our linguistics nor the variety of ways it may develop is, of course, exhausted by the group of problems outlined above. The editors deem it important and desirable that linguists in the U.S.S.R. and the foreign readers of our journal comment on this article.

II. VARIOUS TYPES OF HOMONYMS AND METHODS OF DISTINGUISHING THEM IN MACHINE TRANSLATION

(Based on English, German, Russian, Chinese, and Japanese materials)

Pages 97-101

S. S. Belokrinitskaya,
A. A. Zvonov, M. B. Yefimov,
T. M. Nikolayeva,
and G. A. Tarasova

The number and kind of lexical and grammatical homonyms in machine translation is closely and directly connected with the type of dictionary used in the particular system of machine translation. In the algorithms that we have examined the dictionaries are constructed as follows. The English dictionary includes nouns in the singular number, common case, adjectives and adverbs in the positive degree (suppletive degrees of comparison are registered as independent words), verbs in the basic form as in ordinary dictionaries. The Chinese dictionary includes words in their basic lexical form. The German, Russian, and Japanese dictionaries contain stems.

The routines for distinguishing homonyms, which occupy a special place in the system of lexical analysis of a machine translation algorithm, do not consider the cases of lexical homonymy or cases of grammatical homonymy that are handled in the grammatical analysis routines. The set of rules for distinguishing homonyms, described in this article are based on an analysis of lexical and grammatical homonymy.

In describing the types we shall follow this classification: (1) homonyms recognizable out of context on the basis on endings alone, i.e., homostems, and (2) homonyms distinguishable only by analysis of the context, i.e., homoforms. Homonyms of the first type are called "lexical," homonyms of the second type "contextual."

Lexical homonyms

Among the lexical homonyms we distinguish: (a) homonyms belonging to different parts of speech, and (b) homonyms within a single part of speech.

(a) Stems of nouns and verbs are numerous among the homonyms of different parts of speech in Russian and German, e.g., Russian kos-(kos), kosit' - kosa [to mow-scythe], zamen-(zamen-), zamenit'-zamena [to substitute-substitution]; German Frag-(frag-), die Frage - fragen, band-(Band), binden - der Band. Homonyms of this class are distinguished by checking for a certain type of ending characteristic of one part of speech and not of another.

Lexical "adjective-noun" homonymy in Russian (tsel-tselyy/tsel' [whole/goal]; chast-chastnyy/chast' [private/part] etc.) and "adjective-verb" homonymy in German (gleich-gleichen) are distinguished in similar manner.

The following classes of lexical homonyms are distinguished in machine translation from Japanese into Russian: (1) "noun-adjective"; (2) "noun-verb"; (3) "noun-adjective-verb"; (4) "adjective-adverb." The classes thus identified are not found in traditional Japanese grammar, which assumes the presence of two main parts of speech: taigen (nouns) and yogen (predicates).

A special place is accorded to the differentiation of homonyms in which the stem of the first word and the entire second word coincide.

(b) Homonymy of stems within a single part of speech is likewise distinguished by analysis of the specific endings that are possible for one grammatical type and not for another. This applies, in particular, to homonymy of nouns belonging to different declensions and coinciding in the dictionary in a single lexical stem in Russian (os-/osa-oc' [wasp-axis]) or to aspect homonymy within the same verb (prida-/pridam-pridayu [I shall give-I am giving]).

The actual work of differentiating all the types of homonyms described above is done on the basis of special lexical symbols with which each potentially homonymic stem is initially furnished. The differentiation process in the general system of machine translation algorithms takes place right after the particular stem is found in the dictionary, i.e., after the input text is handled by the "Rejection of Endings" routine.

Contextual homonyms

By contextual homonymy we understand primarily conversion homonymy. Contextual homonymy is determined and differentiated chiefly by analysis of individual word combinations. The purpose of this analysis, it should be borne in mind, is not to obtain a translation of homonymic forms, but to generate a sign of the part of speech for the word undergoing analysis. Five main classes of conversion homonyms have been identified in English: "verb-noun" (class 1), "noun-adjective" (class 2), "verb-adjective" (class 3), "adjective-adverb" (class 4), and "preposition-adverb" (class 5). The morphological poverty of English words is responsible for the little help available to us for morphological analysis in connection with differentiation of homonyms. That is why syntactic analysis is so important along with some lexical analysis of the context. The rules of homonym differentiation are based on the principles of morphological and syntactic analysis* [*Cf. T. N. Moloshnaya, "Problems in the Differentiation of Homonyms in Machine Translation of English into Russian,"

in the collection Problemy kibernetiki [Problems in Cybernetics], No. 1, Moscow, 1958).] Combined with simultaneous use at times of contextual lexical analysis, When checked on a substantial amount of new material, the results of the routine were quite satisfactory.

Syntactic analysis in homonym differentiation depends on the identification of grammatical combinations characteristic of words in one grammatical class and atypical or impossible for words in another class. It may be very simple. For example, a word can be classified as a noun in differentiating "verb-adjective" homonyms merely by determining whether it is preceded by a $\not$ * [*Here and elsewhere the symbol $\not$ means that it is possible to apply in the given case one of the rules of omission that in checking permit the skipping of words and combinations of certain grammatical types (e.g., adverb, particle, parenthetical word, homogeneous member, etc.).] or definite article "the," indefinite article "a," preposition, numeral, or adjective.

If, however, it turns out during the analysis process that none of the above-mentioned words comes before the given item, an additional check is made for words standing directly after our word. If our word is followed by a formula not preceding a noun, conjunction connecting homogeneous members of a sentence (the so-called "homogeneous" conjunction), it likewise will receive a noun sign, e.g., "...and hence again by integration a third approximation (Figure 1, curve C) is derived."

To determine the part of speech of a homonym, it is sometimes necessary to examine carefully not only the words directly adjacent to the item, but to analyze the wider context grammatically.

Complex homonyms of the "noun-verb-adjective" class (e.g., "check," "end," "group," "limit," etc.) are considered in two successive stages in accordance with the aforementioned principles of analysis. These homonyms are given a special symbol in the dictionary and are first directed to the section containing homonyms of the "verb-noun" class. If the process of analysis reveals that the given word is neither a verb nor a noun, it is sent back to the "verb-adjective" class where it obtains the appropriate grammatical information.

Thus far we have discussed only conversion homonyms having a paradigm. Our special group of homonyms, in addition to the cases already noted, contains four rules for differentiating homonyms of the "preposition-adverb" class.

The difficulty in solving the problem of conversion in machine translation from Chinese into Russian is intensified by the almost total lack of morphological structure of Chinese words. The main elements of value in working out the conversion are strict word order, syntactic combinability of certain classes

of words, and auxiliary words. There is a special group of rules for each conversion class which produce a sign indicative of the part of speech to which the words belong.

In distinguishing "noun-verb" homonyms, besides analyzing the possible suffix structure of the particular class of words, we take into account the ability of Chinese nouns and verbs to combine with formants of the past tense and perfective aspect (tsen, tsentszin, i, itszin), the formant so, modal verbs, copulative verbs, negations, adverbs, suffixes of unity, adjectives, prepositions, postpositions, auxiliary words, etc.

Let us see how this works in several examples: Tszin' taolun' gen gyanfa dy tsinkuan, "Let us now consider a more general case"; Dinli yuyshy chzhenmin, "Then the theorem is proved"; Tsy fanchen yu vey i tsze, "The given equation has a single solution"; In'vey chzhenmin shi shifen' tszyan'dan', "Since the proof is trivial."

The "noun-adjective" class is subdivided by us into two subclasses: "noun-relative adjective" and "noun-qualitative adjective." The need of this subdivision arises from the impossibility of analyzing simultaneously types of words differing lexically and grammatically. Words of this type found before nouns are given the sign "relative adjective," otherwise "noun."

Words belonging to the "noun-qualitative adjective" class (tungou - "isomorphism-isomorphic"; dan'chun'tungou - "homomorphism-homomorphic," etc) are analyzed with account taken of the fact that, as shown by observations of the combinability of such pairs with other parts of speech, words of this type functioning as a predicate stand at the end of a clause and have in front of them two or more nouns joined by the conjunctions yuy, gen', khe, or tun. It will be noted that grammatical analysis of this class of words is simplified by generation of the sign "relative adjective" in cases where they come before the noun.

Solution of the conversion of the "adjective-adverb" class presents no special difficulties since the only requirement is to ascertain whether such a word comes before a verb (or it contains the formant di). If so, it is an adverb.

All Chinese modal verbs may as a matter of fact appear in certain contexts as modal adverbs (as in the preceding cases, we start from the requirements of translation into Russian). The main criterion for generating the sign "modal verb" for this class of words is the presence in the sentence of a subject expressed by an animate noun.

Analysis routines have been prepared for each of these conversion classes. In those cases where there is ambiguity along with conversion homonymy the words are given the sign "polysemantic" and analyzed in accordance with the polysemy routine.

The unusual method of differentiating homonyms in Japanese is due to the way the sentences are broken down into words in machine translation. Since the writing is in the form of an alternation of characters and syllabic signs (kana), we first isolate the words from the period to the nearest character, from one combination of characters to another or to any mark of punctuation or formula. Thus, in one case sekibun hoteisiki ("integral equation") is isolated as a single conventional word, in an other yomikata ("reading") as two words.

In German differentiation of sets of homonyms involving nouns exploits the peculiarity of orthography wherein the noun is always written with a capital letter. We can thus say that the process is based on the graphic-syntactic principle. Analysis of the words included in the "noun-verb" (Beweis-beweisen, Gebrauch-gebrauchen, etc) and "noun-adjective" (Komplex - komplex, Parallele-parallele, etc) classes begins with a check of the spelling for a capital letter. A negative answer yields a simple solution: the particular word is not a noun, but a verb or adjective. If the answer is positive, the preceding word is checked for a "period," which sometimes also yields a solution. If the preceding word is not a period, the item being analyzed is a noun.

Syntactic analysis in connection with the "noun-verb" class of homonyms is based on ascertaining the difference in types of sentences shown by the position of 2 verb-predicate or verbal part of a compound predicate and by the relatively fixed word order of German sentences in general, especially in scientific texts. Another feature of the process of distinguishing "verb-noun" homonyms is the need of subjecting all verbs without exception to processing by the appropriate routine since in German it is both theoretically and practically possible to substantivize any verb in the infinitive form.

Differentiation of homonymy of invariable words ("preposition-particle") is based on the fact that these two categories of words belong to different syntactic groups. Prepositions serve to express syntactic relations between words in the noun group, whereas particles, as a rule, form part of the verb group. Therefore, a preposition is found by checking the following word for a noun (or formula or number which can perform the syntactic functions of a noun). For example, it is easy to determine that in the following sentence the first zu is a preposition and the second zu a particle: Um zu diesen Zahlenwerten zu gelangen, sind erhebliche konstruktive Arbeiten notwendig gewesen [In order to obtain these values, considerable constructive work was required]. Conversion homonymy in Russian is accidental and very rare.

The most difficult and pressing problem is to differentiate the homonymy existing between different noun declensions. Let us consider, for example, the paradigms of the singular number of the first declension noun matematik [mathematician] and of the second declension noun matematika [mathematics]:

Nom.	matematik	matematika
Gen.	matematika	matematiki
Dat.	matematiku	matematike
Acc.	matematika	matematiku
Inst.	matematikom	matematikoy
Prep.	matematike	matematike

It is evident that only the forms of the instrumental of both declensions* [*The nominative of matematik coincides with the genitive plural of matematika while the genitive of matematika coincides with the nominative plural of matematik.] are not homonymic.

A special place in contextual homonymy is accorded to the "adjective-adverb-category of state" type characteristic of words like khorosho [well], plokho [badly], etc. In the cases mentioned above it was a question of the coincidence of forms of two different words, but here, on the contrary, one initial amorphous word depending on its syntactic function may change its morphological classification. Thus, the same word khorosho may be an adverb (khorosho sdelano [well done]), category of state (mne khorosho [(lit.) to me is well]), or nominal part of a predicate. The syntactic factors here prevail over the morphological factors, i.e., the syntactic function of the word in a sentence determines the part of speech to which it belongs, and not vice versa. The phenomenon of lexical-grammatical homonymy that is so widespread in all languages is one of the least studied problems in lexicology and grammar.

It is our view that the classification of homonyms by types and the differentiation of homonyms of different classes possess on the whole significance not only for applied, but also for general linguistics, although we considered it our main task to find a solution for the problem of homonymy in connection with machine translation. In most cases the methods described in this article yielded positive results.

III. CONFERENCE ON MACHINE TRANSLATION IN SWEDEN

Page 154 Unsigned article

At the initiative of G. Birnbaum, a lecturer in the University of Stockholm, a conference was held in Stockholm on December 1, 1959 to discuss the organization of research on machine translation in Sweden. Birnbaum and engineer U. Dopping presented papers at the conference in which linguists, mathematicians of the Royal Technical Institute in Stockholm, communication engineers, and others took part. Birnbaum set forth the general prerequisites for scientific and practical work on machine translation in Sweden. Dopping reported on computers in the country that might also be used for machine translation and on the machines ordered for the new few years. Birnbaum's proposal to organize a team of linguists, engineers, mathematicians, etc to work on machine translation was unanimously approved. The group will be temporarily under the jurisdiction of the State Committee on Computers. The Committee will collect all the available literature on the subject. Birnbaum supplied a preliminary bibliography issued in the U.S.S.R. The conference also discussed the main problems to be worked on by the newly organized group. One of the source languages will be Russian.

IV. NEW PUBLICATION ON MACHINE TRANSLATION

Pages 159-160 Unsigned article

In May 1959 the Association for Machine Translation of the First Moscow State Pedagogical Institute of Foreign Languages issued a periodical entitled Mashinnyy Perevod I Prikladnaya Lingvistika [Machine Translation and Applied Linguistics].* [*Subscription orders for the bulletin (cash on delibery) are to be sent to Laboratory of Machine Translation, First Moscow State Pedagogical Institute of Foreign Languages, Metrostroyevskaya, 38, Moscow.] The articles published in the bulletin deal with specific questions in machine translation, construction of algorithms, and general theoretical problems in structural linguistics. The bulletin also describes the activity of machine translation organizations in the U.S.S.R. and the major achievements of foreign research centers. There are reviews of books, articles, and collections appearing abroad and a chronicle of the activities of the Association.

The first number gave an account of the All-Union Conference on Machine Translation held in May 1958. Besides listing the titles of the papers presented, the bulletin included the full text of the reports of V. Yu. Rozentsveyg (on the work of the theoretical section of the conference) and of V. A. Uspenskiy (on the work of the algorithm section).

The second and third numbers contained a number of interesting articles. O. S. Kulagina familiarized her readers with various types of operators used in machine translation. Yu. K. Lekomtsev tried to analyze entire sentences as non-elementary (derivative) signs differing in degree of realization of some regular, full scheme. V. A. Nikonov cited the results of a statistical investigation of the interrelation of cases in modern Russian, where the genitive occurs with the greatest frequency. R. M. Frumkin discussed methods of compiling frequency dictionaries. Two papers dealt with machine analysis of Georgian and Armenian affixation. T. M. Nikolayeva examined the problem of automatic determination of aspect in the Russian verb from the context. N. N. Leont'yeva described ways of handling the conjunction i [and] in machine translation.

These numbers of the bulletin also contained reviews and surveys: I. A. Mel'chuk - the original grammatical dictionary of M. Gabor (Hungary) and the All-Union Conference on Mathematical Linguistics held in Leningrad April 15-21, 1959. A short note by B. A. and V. A. Uspenskiy described a statistical study of English syntagmas in which computers of the National Bureau of Standards (U.S.) were used. A. A. Babintsev discussed

the operation of the Japanese translation machine the Yamato. The bulletin also chronicles in detail the work of the Association and gives a bibliography of books received by the Laboratory of Machine Translation of the First Moscow State Pedagogical Institute of Foreign Languages.

The fourth number will include articles on experiments in machine translation from French into Russian (O. S. Kulagina), grammar in the intermediary language (I. A. Mel'chuk), the text of a report by N. D. Andreyev, the present status of work on machine translation in the U.S.S.R. (V. V. Ivanov and I. A. Mel'chuk - the paper was presented at the conference on computer technology held in Moscow during November 1959).

5214

- END -